

AFOSR 69-0743 TR

AD 684626

TRIAL: An Information Retrieval System
for Creating, Maintaining, Indexing,
and Retrieving from Files of
Textual Information

USER'S MANUAL

1. This document has been approved for public release and sale; its distribution is unlimited.

Lorraine Borman
Donald Dillaman
Vogelback Computing Center
Northwestern University
Evanston, Illinois
October 1968
Revised December 1968
NUCC118

DDC
RECEIVED
MAR 26 1969
B

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

Supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR 68-1593, entitled "Man-Machine Interaction in Information Retrieval Systems". This support contributed to the completion of TRIAL for the CDC 6400 and is intended to aid the development of a remote terminal retrieval system.

TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| Program Description. | 1 |
| Flow Diagram | 2 |
| Input Data Preparation | 3 |
| File Structure. | 3 |
| Input Data Format | 4 |
| Levels of Information | 8 |
| Creating a Master File (the EDIT routines) | 13 |
| EDIT Control Cards. | 14 |
| Sample Job Deck Structure 1 | 21 |
| Sample Job Deck Structure 2 (card input). | 22 |
| Sample Job Deck Structure 3 (tape input). | 23 |
| Sample Job Deck Structure 4 (update). | 24 |
| Search and Retrieval Methods | 25 |
| A. SEARCH Program | 25 |
| The Search Command. | 25 |
| Title Card. | 26 |
| What to Search. | 26 |
| Use of Boolean Operators. | 27 |
| Weighting of Search Terms | 28 |
| Control Cards | 30 |
| Sample Job Deck Structure | 31 |
| B. INDEX Program | 32 |
| KWIC (keyword-in-context) Indexing. | 32 |
| KWOC (keyword-out-of-context) Indexing. | 35 |
| Control Cards | 40 |
| PRINT Options | 43 |
| Sample Job Deck Structure | 46 |
| References | |
| Appendices | |

LIST OF FIGURES

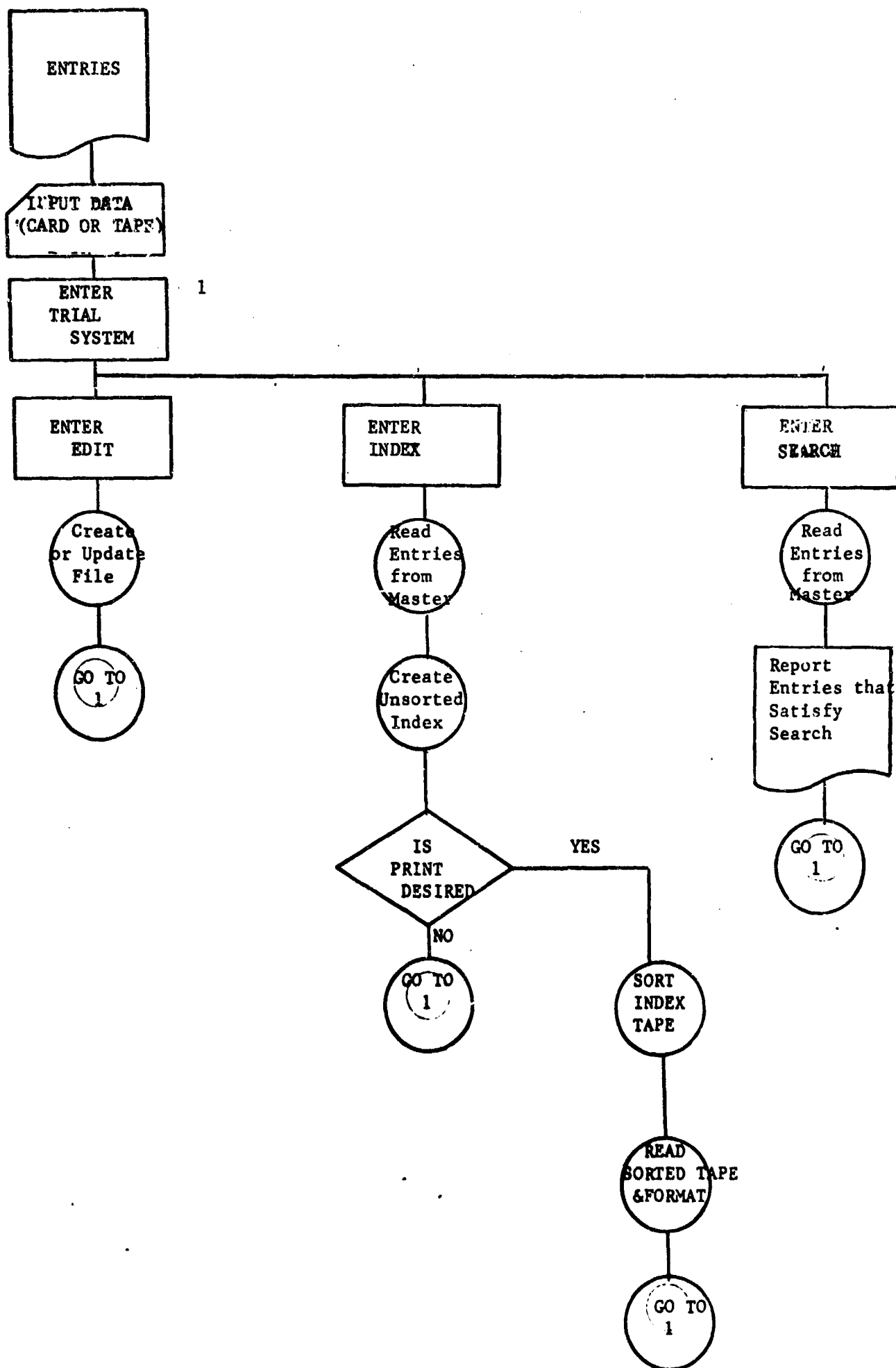
| | <u>Page</u> |
|--|-------------|
| 1. Keypunch Instructions for TRIAL Bibliographic Applications. . | 5 |
| 2. Bibliographic Coding Form | 6 |
| 3. Sample Format for M-R Codebooks | 7 |
| 4a. Sample Input Using Repeating Groups | 9, 10 |
| 4b. Search Command of Repeating Group Information | 11 |
| 4c. Output Illustrating Repeating Group Concept | 11 |
| 5. Sample Deck to Create a Master File from Card Input | 22 |
| 6. Sample Deck to Create a Master File from Tape Input | 23 |
| 7. Sample Deck to Update a File. | 24 |
| 8. Sample Search Command | 27 |
| 9. Output from Search Command. | 29 |
| 10. Sample Deck to Search and Retrieve Items from an Existing Master File | 31 |
| 11. KWIC Output, Wraparound | 33 |
| 12. KWIC Output, using NOWRAP Feature | 34 |
| 13. Title Index to Vogelback Computing Center Library - KWOC Format. | 37 |
| 14. KWOC Index by Author. | 38 |
| 15. *AUTHOR(short-form) Index | 39 |
| 16. Sample Deck to Produce Author and Title Indexes in KWOC Format | 46 |

Program Description

TRIAL is an information processing system that will perform editing, indexing, and retrieval of textual and certain types of numeric information. The system allows for the creation and maintenance of a master file (EDIT), indexing on words designated as "key" words, alternatively, on every word in the text, excluding those common terms that are user supplied as a "stop" word list (INDEX), and computer retrieval and printout of entries that satisfy a user search command (SEARCH). The system is designed such that any one or any combination of the above features can be achieved through one computer run with proper control cards.

The system is sufficiently flexible to handle diverse applications in information retrieval. TRIAL has been successfully used at Northwestern for a "Selective Dissemination of Information" (SDI) system which automatically notifies social scientists of new journal articles that appear to fit their personal interests, to retrieve information from data files where the description of the study and the questions or variables used were in machine-readable form, and to select students for overseas work by retrieving from personnel files those individuals whose background satisfied the requirements of the project. The system is especially adaptable to large masses of bibliographic data where either selective bibliographies or various forms of indexes are desired.

*TRIAL was originally written in machine language for the IBM 709 (1964) and rewritten for the CDC 3400 (1966) by William H. Tetzlaff of Northwestern University under the supervision of Professor Kenneth Janda of the Political Science Department. See References #9, 10, 12.



FLOW DIAGRAM OF INFORMATION THROUGH TRIAL SYSTEM

INPUT DATA PREPARATION

The TRIAL system is a series of user-oriented FORTRAN routines designed to facilitate the computer retrieval and manipulation of textual information. Design specifications have emphasized ease of use, modularity, effectiveness of retrieval, and a realistic approach to the development of a cost-conscious system.

File Structure

The TRIAL file structure allows for a maximum of nine levels of record definition within an entry; an entry is considered to be all of the information for one unit of analysis. It may be ten cards defining author, title and abstract for an entry in a bibliographic file, or 500 cards describing a FORTRAN computer program: input/output specifications, error conditions, size limitations, etc. Individual applications as designated by the research design may utilize from one to all nine distinct levels of information. Any level may have new information added to it, or, if original research needs change, new levels may be added to the existing file.

The first step of file creation is definition of TRIAL structure. Each information level is defined by the user so as to meet his needs. Although the programs do not expect specific forms of input within any of the levels (with the exception of level one), it must be stressed that consistency over records in a file is of utmost importance. A standard bibliographic file might include:

- Level 1: Author
- Level 2: Title
- Level 3: Source and date of publication
- Level 4: Library call number
- Level 5: Descriptors: textual or coded
- Level 6: Abstract

The first three levels of information would be considered as essential before entering the file; items 4, 5, and/or 6 might be originally available and if so, would be entered as original input. Any entry, in this case, bibliographic reference,

could contain any or all of the six defined sets of information. Information within any level could be added to an entry by subsequent update of the file. In addition, levels 7, 8, and 9 might be utilized also in the future as research needs and goals become more defined.

Input Data Format

The input format utilizes 60 characters of free-field information per line, a one character numeric identifier of the level of record definition in column 75, and nineteen characters of optional reference code. This 60-character text per line format is generally compatible with IBM KWIC, the Inter-University Consortium for Political Research codebook files, and the recommendations established by the Council for Social Science Data Archives for machine-readable codebooks. Key punching of data begins after research needs and file definitions have been established. One method of preparing bibliographic entries is shown in Figure 1, Key punch Instructions for TRIAL Bibliographic Applications, and Figure 2, TRIAL Bibliographic Coding Form. A file of machine-readable codebooks to survey research studies might follow the coding instructions illustrated in Figure 3, Sample Format for Machine-Readable Codebooks-Intersocietal Information Center.

Note: Input may be any alphanumeric characters. The only restriction is that a line (card) may not begin with an asterisk.

Vogelback Computing Center Library
Bibliographic Coding Form
Instruction Sheet for Key punching

Author cards: Level 1

Last name, first name (or initials)

Punctuation only if desired.

For individual authors, 18 columns per entry:

Author 1 col. 1-18

Author 2 col. 21-38

Author 3 col. 41-58

Author 4 col. 1-18 card #2 of type 1

If corporate author, punch col. 1...n as needed.

Title cards: Level 2

Titles can be punched as desired, free field, col. 1-60.

Use as many cards (lines) as needed.

Source cards: Level 3

Col. 1-60 free field. Punch as desired. Do not prepare copy with semi-colons.

Suggested format:

city, publisher, year.

Abstract cards: Level 5

Col. 1-60 free field. Indenting for paragraphs by starting first card in col. 3 or col. 5, or alternating, card 1, col. 1 and succeeding cards, col. 3.

Do not hyphenate words. If a word cannot be completed in col. 60, go on to a new line (card).

It is suggested that abstracts be typed using 60 characters per line so that the keypuncher can exactly follow the copy provided.

DATA CARDS MAY NOT BEGIN IN COLUMN 1 WITH AN ASTERISK.

Figure 1.
Key punch Instructions for TRIAL Bibliographic Applications.

Vogelback Computing Center Library
Bibliographic Coding Form

Lorraine Borman
Northwestern University
April 1968

| | | Level | Accession No. |
|--------------|---|---------------------|------------------|
| Author: | Love, Patricia Oldfather, Paula | Col.75 Col.76-80 | |
| Type 1 | col. 1-18 col. 21-38 | 1 | 889 |
| Title: | Programming by Questionnaire - Auxiliary Programs | 2 | |
| Type 2 | col. 1-60 | | |
| Source: | Rand Corporation, Santa Monica California | 3 | |
| Type 3 | col. 1-60 August, 1968 | | |
| Descriptors: | 42 | 4 | |
| Type 4 | | | |
| Abstract: | | 5 | |
| Type 5 | | | |

Programming by Questionnaire is a method by which many computer programs can be produced with a considerable saving of time, effort, and cost. The present Memorandum complements two others in this series by presenting five small computer programs that facilitate the application of Programming by Questionnaire in general, and provide additional analysis of the JSSPC programs.

Figure 2.

STUDY 136 A COMPARATIVE ANALYSIS OF THE RELATIONSHIP
BETWEEN SEVERE ENVIRONMENTS AND SACRED ORIENTATIONS

THIS STUDY REVIEWS THE UNIFORMITIES AMONG THE SACRED ORIENTATIONS OF SOCIETIES WHICH EXIST IN THE MOST SEVERE ENVIRONMENTS ON EARTH. THE TOTAL DATA FOR EACH SOCIETY ARE FORTY-SIX VARIABLES INCLUDED UNDER THE FOLLOWING TOPICS--- ENVIRONMENT AND RELEVANT CHARACTERISTICS, ORGANIZATIONAL STRUCTURE, ALLOCATION OF RESOURCES, ROLE-DIFFERENTIATION, IDEATIONAL STRUCTURE, RELIGIOUS ORIENTATION, AND RECENT HISTORY. THERE ARE FOURTEEN SOCIETIES WITH ONE CARD OF DATA PER SOCIETY.

IN 1963, THE DATA WERE COLLECTED BY BERNARD BECK AND WERE SUBMITTED TO THE DEPARTMENT OF SOCIOLOGY AND ANTHROPOLOGY AT PRINCETON UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIRMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY. THE RESEARCH WAS SUPPORTED BY THE NATIONAL SCIENCE FOUNDATION. ALL INFORMATION WAS GAINED THROUGH LIBRARY RESEARCH.

THERE ARE NO PLUS OR MINUS SIGNS PUNCHED IN THE DATA.
ALL ZERO-S ARE CODED TO INDICATE INSUFFICIENT EVIDENCE.

THE DATA HAS NEVER BEEN PUBLISHED EITHER IN TOTAL OR IN PART.

IDENTIFYING INFORMATION

C. 1-2 CONTAINS THE IDENTIFICATION CODE OF EACH SOCIETY

COL. 3 BLANK

COL. 4-9 CONTAINS THE NAME OF THE SOCIETY CORRESPONDING
TO THE IDENTIFICATION CODE

- 01 ESKIMO
02 YUKAGH
03 URA
04 AHH CA
05 BUSH
06 SHOSU
07 PYG MIE
08 NEGHIT
09 ALGONK
10 AHH TS
11 AHH AL
12 SIPINU
13 VEDDAS
14 LACANDU

COL. 10 BLANK

COL. 11 POPULATION DENSITY (IN PERSONS PER SQUARE MILE)

0. INSUFFICIENT EVIDENCE
1. 1/1 TO 1/5
2. 1/5 TO 1/9
3. 1/13 TO 1/16
4. 1/13 TO 1/35
5. AS LOW AS 1/100
6. AS LOW AS 1/250

| | Level No. | Study No. |
|--|-----------|-----------|
| Title - - - - | 1 | 136 |
| | 1 | 136 |
| | 1 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| Description of - - - Contents | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| | 2 | 136 |
| Data collection, sponsorship, technical information, sources | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| | 3 | 136 |
| Publications reporting -- data | 4 | 136 |
| | 4 | 136 |
| | 4 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| Information about data and units -- of analysis | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| | 5 | 136 |
| Variable | 117 | 136 |
| | 117 | 136 |
| Response Category | 118 | 136 |
| | 118 | 136 |
| | 118 | 136 |
| Card col. of variable | 118 | 136 |
| | 118 | 136 |
| | 118 | 136 |

Figure 3.

Sample Format for Machine-Readable Codebooks - Intersocietal Information Center

Level One Information

Level one information is normally used for authors in a bibliographic file and is processed by the system in a slightly different method than other records. Up to three individual authors may be punched in each author card, using columns 1-18, 21-38, and 41-58. The authors last name is punched first, followed by his initials (or first name). Punctuation may be included.

A corporate author is signified by the presence of a character in either column 19 or 20. A name such as ASSOCIATION FOR COMPUTING MACHINERY would be treated as a corporate author and for certain indexing purposes would be considered as a single entity; in all other instances, i.e., keyword indexes or searching procedures, each word would be considered as an entity.

Levels Two Through Six

Information levels two through six are recorded as free field using sixty characters per line with the restriction that a word may not continue from one line to the next. Hyphenation operates to treat the hyphenated characters as one word.

Levels Seven Through Nine

Levels 7 through 9 are, for indexing purposes, handled in the same manner as preceding levels (2-6) but are considered as "repeating group" information by the search and retrieval routines. This data may be the codebook of a study on underdeveloped nations where the level 7 information would be the variable being studied, level 8 the coding categories, and level 9 source of coded information. In other applications the repeating group concept might be considered as "propositions" and "elaboration". Searching levels 7-9 would retrieve the identification levels of an entry together with only those repeating groups which satisfy the search command. Figures 4a,b,c show repeating group input (4a) and the output (4c) produced by a search (4b) of level 7 information. In this example the file to be interrogated contains information on 200 studies of a cross-national nature.

1 136
1 136

| | |
|---|------------|
| 1 | 136 |
| 2 | <u>136</u> |
| 2 | 136 |
| 2 | <u>136</u> |
| 2 | 136 |
| 2 | <u>136</u> |
| 2 | 136 |
| 2 | <u>136</u> |
| 2 | 136 |

[illegible]

3 136
3 136
3 136

| | |
|---|-----|
| 3 | 136 |
| 4 | 136 |

4 136
4 136
5 136

| | |
|---|-----|
| 5 | 136 |
| 5 | 136 |
| 5 | 136 |

| | |
|---|-----|
| 5 | 136 |
| 5 | 136 |
| 5 | 136 |

5 136
5 136
5 136

5 136
5 136

5 136
5 136

5 136
5 136

| | |
|---|-----|
| 3 | 136 |
| E | 126 |

| | |
|---|-----|
| 5 | 136 |
| 5 | 136 |
| 5 | 136 |

5 136
5 136

5 136
5 136

5 136
5 136

136

7 136
7 136

136
136

136

136

136

136
136

136
136

136

1

| | | | |
|---------|---|-----|-----|
| COL. 13 | POPULATION DECLINE | 137 | 136 |
| | 0. INSUFFICIENT EVIDENCE | 138 | 136 |
| | 1. EXTINCTION | 138 | 136 |
| | 2. VERY GREAT DECLINE SINCE OUTSIDE CONTACT | 138 | 136 |
| | 3. SOME DECLINE | 138 | 136 |
| | 4. NO DECLINE | 138 | 136 |
| | 5. VARIABLE EFFECT | 138 | 136 |
| | : | | |
| | : | | |
| | : | | |
| | : | | |
| | : | | |
| COL. 40 | LEADERS | 407 | 136 |
| | 1. FATHER OF FAMILY AND/OR THE EXTENDED FAMILY | 408 | 136 |
| | 2. INFORMAL AND TEMPORARY | 408 | 136 |
| | 3. HEADMAN AND SPECIFIC LEADERS FOR SPECIFIC AREAS | 408 | 136 |
| | 4. OLDER MALES, AND BAND LEADERS | 408 | 136 |
| | 5. INITIATED MEN AND HEADMAN | 408 | 136 |
| | 6. MATURE MEN, AND HEADMAN | 408 | 136 |
| | 7. BAND HEADMAN | 408 | 136 |
| | 8. BAND LEADER AND EXTENDED FAMILY LEADER | 408 | 136 |
| | 9. CLAN HEADMAN AND COUNCIL OF OLDER MEN | 408 | 136 |
| | : | 408 | 136 |
| COL. 41 | FOREIGN RELATIONS | 417 | 136 |
| | 1. LIMITED | 418 | 136 |
| | 2. MINIMAL AND OFTEN HOSTILE | 418 | 136 |
| | 3. MINAMAL BUT GENERALLY PEACEFUL | 418 | 136 |
| | 4. AVOID INTER-GROUP CONTACT EXCEPT FOR TRADING AND GIFT EXCHANGE | 418 | 136 |
| | 5. DEFENSIVE AGAINST INFRINGEMENT | 418 | 136 |
| | 6. SUSPICION, BUT NO HOSTILITY | 418 | 136 |
| | 7. HARMONIOUS INTER-GROUP CONTACTS WHEN BASED ON KINSHIP PRINCIPLES | 418 | 136 |
| | 9. REGULAR AND WELL-ORDERED | 418 | 136 |
| | : | 418 | 136 |
| COL. 42 | FREQUENCY OF WARFARE | 427 | 136 |
| | 1. UNKNOWN AT PRESENT | 428 | 136 |
| | 2. VIRTUALLY UNKNOWN | 428 | 136 |
| | 3. ONLY OCCASIONALLY | 428 | 136 |
| | 4. UNKNOWN IN MOST AREAS | 428 | 136 |
| | 5. FREQUENT OCCURRENCE | 428 | 136 |
| | 6. CONSTANT, TAKES HEAVY TOLLS | 428 | 136 |
| | : | 428 | 136 |
| | : | | |
| | : | | |
| | : | | |

Figure 4 (a) cont. (open dotted areas indicate intervening data)

CROSS NATIONAL SEARCH FOR PROPOSITIONS,
 (
 (FOREIGN*RELATIONS) .OR. (DIPLOMATIC*RELATIONS) = 100,
)\$

Figure 4 (b) Search of Level 7 Information.

CROSS NATIONAL SEARCH FOR PROPOSITIONS

STUDY 136 A COMPARATIVE ANALYSIS OF THE RELATIONSHIP
 BETWEEN SEVERE ENVIRONMENTS AND SACRED ORIENTATIONS

THIS STUDY REVIEWS THE UNIFORMITIES AMONG THE SACRED
 ORIENTATIONS OF SOCIETIES WHICH EXIST IN THE MOST SEVERE
 ENVIRONMENTS ON EARTH. THE TOTAL DATA FOR EACH SOCIETY ARE
 FORTY-SIX VARIABLES INCLUDED UNDER THE FOLLOWING TOPICS---
 ENVIRONMENT AND RELEVANT CHARACTERISTICS, ORGANIZATIONAL
 STRUCTURE, ALLOCATION OF RESOURCES, ROLE-DIFFERENTIATION,
 IDEATIONAL STRUCTURE, RELIGIOUS ORIENTATION, AND RECENT
 HISTORY. THERE ARE FOURTEEN SOCIETIES WITH ONE CARD OF DATA
 PER SOCIETY.

| Printout of only that repeating group that satisfied the search command. | COL. 41 FOREIGN RELATIONS |
|---|--|
| | 1. LIMITED |
| | 2. MINIMAL AND OFTEN HOSTILE |
| | 3. MINAMAL BUT GENERALLY PEACEFUL |
| | 4. AVOID INTER-GROUP CONTACT EXCEPT FOR TRADING AND GIFT EXCHANGE |
| | 5. DEFENSIVE AGAINST INFRINGEMENT |
| | 6. SUSPICION, BUT NO HOSTILITY |
| | 7. HARMONIOUS INTER-GROUP CONTACTS WHEN BASED ON KINSHIP PRINCIPLES |
| | 9. REGULAR AND WELL-ORDERED |
| | |
| PHRASES.... FOREIGN | * |
| | RELATIONS |

Figure 4 (c) Output p. oduced from search command shown in Figure 4 (b) above.

The user might want to retrieve only those studies containing variables on "gross national product in Nigeria." A level 7 search of the file would retrieve and print only the names of those studies in the file containing data on Nigeria and GNP, their location in the data file, and the coding procedure used, i.e., in dollars, in pounds, etc. Since the user is not interested in other variables that might be part of the study, the use of the repeating group search enables him to obtain a quick response to his search command but one which satisfies his needs. Another individual might be interested in variable construction; he would want a printout of a total codebook and would, therefore, use a generic search command, possibly the name of a known study, so as to retrieve an entry in its entirety. A level 7 search, however, is used to retrieve specific selections of information from the body of entry.

TRIAL PROCESSING: CREATING A MASTER FILE

TRIAL consists of six programs within an "overlay"* structure: TRIAL, EDIT, SEARCH, INDEX, SORTER and PRINT. TRIAL serves as the Executive, issuing calls to the other programs as needed. Processing of new input data is the function of the EDIT routines.

In many research situations where large amounts of data are involved, it is usually necessary to place all of this information on a "master file" tape for subsequent processing. The EDIT program consists of ten routines which generate this file.

The EDIT programs analyze instruction cards and set appropriate switches. New input to EDIT is internally reformatted for speed in processing and for ease of update by subroutine CREATE. A header word is generated for each entry containing (1) a reference code as user specified, (2) the date of entry into the file, (3) a serial number for updating, and (4) each line of an entry is numbered by tens for subsequent alteration within an entry. Each record is handled individually with information levels one through six analyzed before any of the repeating group levels are scanned. The WRITEMS, PUT and PUT1 routines write the master file and all other files as specified by the user instructions. Output records consist of 504 words per block. Error checking of new and update input is performed according to user option. (These functions are currently being revised; a memo will be distributed to TRIAL users at the time of completion.)

Many circumstances require frequent changes or additions to the master file as errors are detected or as new information is collected. The procedures connected with this task are known as updating the master file. The

*overlay: The technique of repeatedly using the same blocks of internal storage during different stages of a problem, e.g., when one routine is no longer needed in internal storage, another routine can replace all or part of that storage. The overlay concept thus permits the breaking of a large program into segments which can be used as required to implement problem solution.

TRIAL system allows for the deletion of entries within a file (DELETE), replacement of one entry by one or more entries (REPLACE), insertion of entries after a given entry (AFTER), and alteration within an existing entry (ALTER). Any number of update commands are acceptable with one call to the EDIT program.

All of these update commands allow the user to modify the existing master file and process either indexing and/or searching instructions in one computer run.

All update processes are performed sequentially. Although the use of disk files allows updating on a scratch file with the updated file rewritten on the original master, good programming practice, for obvious reasons, does not recommend this type of usage. Especially in the case of large files, where original information may not be readily accessible, the use of input and output magnetic tape files is recommended. The EDIT master file can be processed immediately as part of the original job or, if written on a magnetic tape, can be saved for future processing.

EDIT Control Cards

*EDIT

This card calls in the various editing routines and signifies that new data or update information are to be inputted to a master file.

Instruction Card

Various files may be assigned at the option of the user. In addition, he may also request options which will 1) produce listings of the input data (LIST), 2) specify the type of reference code to be generated for the file (REFERENCE CODE TYPE = 1), and 3) instruct the system to only process additions to an existing file (ADDITIONS = u).

Instruction Card (cont'd.)

All instructions are typed in free-field format separated by a comma. The last command must be followed by a period. All words enclosed within brackets below are optional.

NEW MASTER = u, (required)

This is a file assignment for the unit number of the master file to be created. If TRIAL is being operated in a disk file system environment, u may be assigned to any file accessible to user programs. Files 1 through 5 are available at Northwestern University.

[OLD] MASTER = ' , (optional)

This command assumes that a magnetic tape file has previously been generated and is to be updated during this computer run. u would equal the logical unit number of the tape drive to be assigned to the existing master tape.

[ALTERNATE] Input = u, (optional)

The use of the INPUT file assignment assumes that data is to be inputted to the system from a magnetic tape rather than cards.

LIST = u, (optional)

The LIST command instructs the EDIT routines to write a file containing a formatted listing of the master file that can be printed off-line. u may equal a logical number of a tape to be assigned or may be equated to the normal output unit; at Vogelback this would be 61.

Instruction Card (cont'd.)

NO MAP, (optional)

A MAP of the master file is automatically generated by the EDIT routines. The MAP indicates the user defined reference code and the internal sequence number assigned by TRIAL to each entry together with the date the entry was added to the master file. The MAP serial number must be referenced if update capabilities are to be utilized. MAP is automatically printed as part of the output file if not specified.

[JUST]ADDITIONS = u,

In many instances a user desires to search or index only additions that have been entered onto the master file. The ADDITIONS command instructs the program to write new entries onto file "u" for subsequent processing. It should be noted that the master file generated will contain all entries. This file assignment is extremely valuable when working with massive amounts of data or in the intermediary stages of file building.

REFERENCE [CODE TYPE] = 1,

Reference codes, such as Defense Documentation Center numbers, or accession numbers, are user defined and may or may not be present in an entry. All reference codes are constructed from the first line of level 1 information of an item. It is suggested, however, that when constructing a large file, a serial number should be included on every card of an entry.

REFERENCE [CODE TYPE] = 1,
(cont'd.)

These code fields may, upon option, be used for indexing and/or search and retrieval.

TRIAL recognizes and processes five types of reference codes:

i = 1 columns 61-72, 76-80
i = 2 columns 1-10, 61-67
i = 3 columns 1-5, 61-67, 76-80
i = 4 columns 61-67, 1-5, 76-80
i = 5 columns 76-80

If the REFERENCE CODE command is not specified, type 1 code is assumed. An example of

the Instruction card to create a master file, reading data from a magnetic tape, and containing a serial number in columns 76-80 to be used as a reference code would be:

NEW MASTER = 1, INPUT = 3, REFERENCE = 5.

Logical units 1 and 3 would be assigned to the physical tapes using local installation system controls.

*DATA (required)

The DATA command signifies the beginning of data to be edited. Instructions for EDIT such as INSERT and the update commands are considered, in this context, to be "data".

Instruction Commands

*INSERT, ENTRY = x, INCREMENT = y.

The INSERT command instructs the EDIT routines to "insert" all entries following the command card onto the master file. It is mandatory when creating a new master file. x is the initial value of the serial number to be

Instruction Commands (cont'd.)

generated by EDIT. Unless specified, it is set at 10. y is the value that the internal serial number will be incremented by for each successive entry. If not specified, increment is set at 10. If a file of 50 entries was to be generated, using *INSERT., their MAP would read:

```
entry 1  0010
entry 2  0020
.
.
entry 50 00500
```

Stepping of the increment value by 10 is especially useful when the possible need to insert entries into specific areas of the master file might arise. One example of this use would be in maintaining alphabetical order within the file. Original input entries could be incremented by tens, or even by 100's. If INCREMENT = 10, 9 additional entries could later be placed between two existing entries. If INCREMENT = 100, 99 additions could be added between two existing records.

To update an existing file

For all update commands, the value of n is the serial number of the entry to be referenced as indicated on the MAP of the existing master file.

*DELETE, ENTRY = n.

This command causes the deletion of the entry that has the serial number matching the number specified.

*REPLACE, ENTRY = n.

REPLACE is used when an existing entry is to be deleted and replaced by one or more entries.

*AFTER, ENTRY = n.

AFTER is used to write m entries after the nth entry specified. If the user desired to maintain an alphabetical order on the file he would use the AFTER command. AFTER is also used to place additional entries after the last record that had been on the old master file.

*ALTER, ENTRY = n.

ALTER allows the user to correct or add information to a record on the file. The data cards to be used in an ALTER operation are punched in TRIAL internal record format

See Appendix 1.

*INSERT, ENTRY = n, INCREMENT = n. INSERT indicates that new data is to be written on the master file as the beginning records. These entries will then be followed by the records that were on the old master file. If the old file had been incremented 10 to 500, entry and increment n should be set at 1 to allow for 9 new insertions to be written without upsetting the sequential ordering of the file. INSERT may also be used in cases to insert a new record or

*INSERT, ENTRY = n, INCREMENT = n. records but it must follow either a DELETE,
(cont'd.)

REPLACE, ALTER, or AFTER function.

*OUTPUT

To list the master tape. The *OUTPUT option allows the user to print the master file at any time subsequent to an EDIT operation. This is primarily to obtain a listing when the LIST option was not used during the EDIT run. If multiple copies are desired, the LIST = n option on the instruction card must be used (see p.15).

*SEQUENCE, ENTRY = n, INCREMENT = n.

The *SEQUENCE command is used when a resequencing of the master file is desired. This is especially helpful when many updates have been performed, possibly using different sequencing methods. (See pp 17-18 for discussion of ENTRY and INCREMENT options). If E and I are not specified, entry will begin at 10 and be incremented by 10's.

*END

END signifies the end of the edit process.

For those jobs where either the initial amount of input data is small or where it is assumed that there will be no need to revise or alter the original data, or in a situation which requires only one pass of the data through the program, the EDIT processes need not be of concern to the user. This one-time job can best be illustrated by the following example:

A research report discussing advances in the information sciences and including a survey of the existing literature has been written which must include a bibliography of cited references in alphabetical order. Input to TRIAL would consist of the following:

```
(Local installation system cards)
*EDIT
NEW MASTER = 1.
*DATA
*INSERT.
entry 1      [ BECKER J HAYES D
                Information Retrieval
                Wiley and Sons, 1964.
entry 2      [ . . . .
entry n      [
*END
*INDEX
*WORDS, AUTHOR, CARDS (1), NULL.
*KWCC, AUTHOR, CARDS (1, 2, 3).
*END
*PRINT
SINGLE COLUMN, CARDS (2, 3).
*END
```

Printout of this job would be photo-ready copy that is amenable to any offset process, or, if only one copy is required of the report, could be inserted directly into the document. By writing the master file to a magnetic tape, this bibliography could be subsequently interrogated.

**TO CREATE A MASTER FILE FROM CARD INPUT AND SAVE IT ON MAGNETIC TAPE FOR
FUTURE PROCESSING:**

VCC System Cards

VLORR,CH1235-9999,CM66000,T1000.

REQUEST(TAPE1) HANG TAPE XXXX (user-supplied tape)

LIBRARY(TRIAL) (Check Library Directory on
Center bulletin board for
name of current version.)

MAP,OFF.

LGO.

7-8-9 End of Record

Tape numbers 1 through 4
may be used. Note that
the number used on the
REQUEST card must match
number used on control
card.

TRIAL System Cards:

Required *EDIT

Control Card (various options)
See pp. 14, 15, 16, 17. . . . NEW MASTER = 1.

Required *DATA

Instruction Card (various
options) See p. 17. *INSERT.

Data Cards

Required *END

6-7-8-9 End of Information

Figure 5. Sample Job Deck Structure

TO UPDATE AN EXISTING MASTER FILE AND PRINT A LISTING OF THE UPDATED MASTER FILE:

VCC System Cards VLORR,CH.....

REQUEST(TAPE1) HANG XXXX (existing OLD master file)

```
REQUEST(TAPE2)  HANG XXY (write updated master
                  (NEW) on this tape)
```

LIBRARY (TRIAL)

LGO.

7-8-9 End of Record

TRIAL System Cards:

Required *EDIT

See pp. 15, 16, 17, 18. . . . OLD MASTER = 1, NEW MASTER = 2, LIST = 61.

Required *DATA

See p. 19. . . . *DELETE,ENTRY=103.

See p: 19. *REPLACE,ENTRY=117.

Data cards for entry which replaces #117

See p. 19. *AFTER, ENTRY=200.

Items to be added to file; will be
written after last entry (#200) of
old master file

Required *END

6-7-8-9 End of Information

Figure 7. Sample Job Deck Structure

Search and Retrieval Methods

Retrieval of information can be obtained by two distinct methods: 1) the use of a search command which establishes specific criteria for item retrieval or 2) indexed output of a KWIC or KWOC type format which generally is then used for manual, or desk, retrieval. In Section A discussion will be limited to the first method, the use of English words and operators to establish search criteria and print the retrieval entries. Section B will describe the indexing and print routines.

A. SEARCH

Any master file created by the EDIT program can be searched. The important features of the SEARCH program are 1) its ability to search textual material for specified words or combinations of related keywords, and 2) its use of the Boolean operators, AND, OR, and NOT, giving the user the ability to specify how these words must appear in combination with each other. If he does not retrieve relevant documents, he can rephrase his command, in much the same manner as the user of a library card catalog. Retrieval from a TRIAL search can be of two forms: a "short form" to allow quick scanning of title and author, or a listing of complete information contained within the entry. In the case of bibliographic files, abstracts, extracts, critical reviews, or even full text may be printed.

The Search Command

A search command is punched on control cards which are read by the Trial system. This command instructs the computer 1) what identifying information is to be printed as a "heading" at the top of each output page, 2) what levels of information of cards are to be searched, and 3) what keywords and logical relations among keywords will satisfy the search. A word is defined as an uninterrupted sequence of 20 or fewer word-forming characters: the letters

A through Z and the digits 0 through 9. Blanks and all other characters are non word-forming characters. (NOTE: The INDEX routines treat a hyphen as word-forming.)

The program considers as a phrase any sequence of words (on records of the same level). Phrases may continue across any type of punctuation characters. Hyphenated words must be searched as if they were phrases, for example, "word-forming" would be searched for as "WORD*FORMING". Because digits 0 through 9 are defined as word-forming characters, particular numbers may be searched for as keywords.

Title Card

A card containing a title or identification label is punched by the user. All instruction cards in SEARCH are virtually free of format restrictions, and this label might begin anywhere on the card and continue for a maximum of 60 characters. The program considers the identification label terminated when it encounters a comma which serves not only to end the identification label but also to instruct the program to look for the levels of information to be searched.

What to Search

Searches can be made on any level of information designated in the search command. Numbers representing the levels to be searched are punched after the comma that terminates the heading. These numbers may be punched in any order, and are separated by commas to improve readability. The level designation is terminated by the first left parenthesis enclosing the keyword portion of the search command.

If only levels 7, 8, or 9 (repeating groups) are designated, the search will be conducted only within records of those levels. If, however, other levels are specified, the computer will search across all specified categories.

Output from a search of repeating group levels will print all level 1-5 information plus only the repeating groups that satisfied the search command(see pp.8-12).

After reading the code number or numbers of the levels to be searched, the machine senses that it is receiving a search command when it reads a left parenthesis. Because any given command may contain more than one combination of keywords arranged in "nests" of parentheses, the program continues to read until the number of right parentheses read equals the number of left ones. It will then evaluate the logic of the command, working outward from the innermost set of parentheses.

Use of Boolean Operators

The power of the search command lies in the use of standard logical operators expressed on the punchcard as follows: ".NOT.," ".OR.," and ".AND." Perhaps the use of the operators can best be conveyed by constructing a sample command for the elites and decision making search. Figure 8 shows the command punched on cards.

```
TEST NO 1A - IIC STUDY CODEBOOKS, 7,8,  
(ELITE .OR. ELITES .OR. LEADERS .OR. LEADERSHIP .OR. DECISION*MAKING .OR.  
DECISION*MAKER .OR. DECISION*MAKERS - 100, )$
```

Figure 8.

This command instructs the program to select only those statements containing both one or more keywords specified within the first nest of parentheses and one or more keywords specified within the second nest. If SEARCH finds any statement of a proposition or finding which satisfies this logical combination of keywords, it will print out the statement and its supporting elaboration after printing the citation of the article and a summary of the study. A portion of the computer output produced in response to the above command is reproduced in Figure 9. The dollar sign after the last right parenthesis on the punchcards in Figure 8 indicates the end of information for a single search instruction. Many different searches can be made during one run on the computer.

Weighting of Search Terms

In addition to retrieving items according to logical combinations of keywords, the SEARCH program provides for "weighting" keywords so that retrieval occurs only if the sum of the weighted keywords in a given entry equals or exceeds a previously stated value. This value, or "hit value", is set presently at 100 (see Figure 8). Retrieval of an item is caused, therefore, when encountering one keyword weighted "100", two keywords weighted "50" each, and so on.

TEST NO 1A - IIC STUDY CODEBOOKS

STUDY 500 FIVE NATION STUDY - UNITED KINGDOM, DECK 2

ALMOND - VERRA

NORC 427

JUNE - JULY, 1959

DECK 02 N=1295

UNITED KINGDOM

53.A. WE ARE ALSO INTERESTED IN HOW WELL KNOWN THE NATIONAL LEADERS OF THE VARIOUS POLITICAL PARTIES ARE IN THIS COUNTRY. COULD YOU NAME THREE LEADERS OF THE CONSERVATIVE PARTY. (IF NECESSARY, TAKE DOWN THE NAMES GIVEN AND CHECK ACCURACY AFTER THE INTERVIEW. MARK THE NUMBER CORRECT FOR EACH PARTY. THEN TALLY TOTAL NUMBER CORRECT AT END OF QUESTION.)

53.B. COULD YOU NAME THREE LEADERS OF THE LABOUR PARTY.

53.C. AND COULD YOU NAME A LEADER OF THE LIBERAL PARTY.

TALLY (CODE TOTAL CORRECT)

495. SEVEN CORRECT

1036. SIX CORRECT

1347. FIVE CORRECT

1309. FOUR CORRECT

1179. THREE CORRECT

1250. TWO CORRECT

110-. ONE CORRECT

45. HERE IS A LIST OF THINGS THAT CHILDREN MAY BE TAUGHT IN SCHOOL. (HAND LIST 10) WHICH WAS STRESSED THE MOST IN YOUR SCHOOL.

R17. HAVE FAITH IN LEADERS

4328. OBEY THE LAW

369. HOW THE GOVERNMENT IS RUN

2847. LOVE YOUR COUNTRY

35-. OTHER

THE FOLLOWING WORDS WERE FOUND IN THE ENTRY

| | WORDS | FREQUENCY |
|-------------------------|---------|-----------|
| (This caused retrieval) | LEADERS | 4 |

Figure 9. Partial output produced from search command shown in Figure 6. The file searched contained 105 codebooks to social science studies; 39 were retrieved.

SEARCH Control Cards

***MASTER** = n (optional - depending
upon type of use)

This instruction is used only when an existing (previously created and saved) file of information is to be used as input to the search routine. n may be any unit from 1 to 5 and must be referenced on a system REQUEST card.

***SEARCH** (required)

To call in the search overlay.

Search command: (required)

Free-field; title, maximum 59 characters; comma; numeric punches indicating levels of information to be searched, separated by and ending with a comma; left parenthesis; search words, phrases, and operators; weight, right parenthesis; \$ to indicate end of command. (See Figure 8, page 27)

***END** (required)

To return control to executive overlay.

TO SEARCH AND RETRIEVE ITEMS FROM AN EXISTING MASTER FILE:

VCC System Cards VLORR,CH.....
 REQUEST(TAPE4) HANG 2044 VCC EDIT MASTER
 LIBRARY(TRIAL)
 LGO.
 7-8-9 End of Record

TRIAL System Cards:

Required *MASTER = 4.
 Required *SEARCH
 Title Card (See SEARCH VCC BIBLIOGRAPHY FILE, 2,5, (
 pages 26,27)
 Search Command (See (NATURAL .AND. LANGUAGE) .OR.
 pp. 25-28) NATURAL*LANGUAGE .OR. TEXT*RETRIEVAL
 .OR. TEXT .AND.
 (SEARCH .OR. SEARCHES .OR. SEARCHING
 .OR. RETRIEVAL) = 100,)\$
 Required *END
 6-7-8-9 End of Information

Figure 10. Sample Job Deck Structure
 -31-

B. INDEX

Keyword indexing by computer is an automatic method for generating alphabetized listings of keywords contained in input material. The INDEX overlay of the TRIAL system contains 14 routines for generating keyword indexes. Input to INDEX is any file that has been created by the EDIT program, either in the same or a previous computer run.

The two most common types of keyword indexing are known as KWIC (Keyword-in-context), and KWOC (Keyword out of context). In KWIC indexing, the computer arranges selected keywords in an alphabetized column surrounded by a few words (maximum length of line=100) of the context in which the keywords occur. In KWOC indexing the selected keywords are printed in a separate column alongside the original context.

Figure 11 reproduces a portion of a KWIC Index to the Midwest Journal of Political Science. The keywords are arranged in alphabetical order to the immediate right of the blank column. The index would be used by scanning down the vertical column of keywords to locate those of special interest. Words on the same line surrounding the keyword contain the context in which it is used. The output page width provides room for 100 characters-spaces to be printed on one line. A title* with no more than 100 characters and spaces prints out in full, although a portion of the title may be "wrapped around" and printed before or after the keyword, depending on the position of the keyword. The "wrap" feature may be eliminated by use of the NOWRAP command (see page 41). This is illustrated in Figure 12. It should be noted that NOWRAP does not print context occurring before the keyword.

*The current version (4) of the KWIC routine automatically prints 18 characters of the author's name (level 1) together with title information (level 2) as one line of output.

(TYPE 2)

| KEYWORDS | | REFERENCE CODE: VOLUME | MONTH | | YEAR |
|---|---|------------------------|---------|--|------|
| | | AUTHOR NAME | | | |
| ABORIGINAL JUDICIAL SELECTION-- | THE WORK OF THE ABA COMMITTEE. | GROSSMAN J | 08-A-64 | | |
| ON THE ABUSES OF POWER IN DEMOCRATIC STATES (THEORY). | | SPITZ D | 01-N-57 | | |
| POLITICAL EXPRESSION UNDER THE MATCH ACT AND THE PROBLEM OF STATUTORY AMBIGUITY(UNIT | | NELSON DM | 02-F-58 | | |
| PARTY ACTIVISM IN A RURAL COUNTY. | | ALTHOFF P | 10-F-66 | | |
| POLITICAL ACTIVISM IN WISCONSIN. | | EPSTEIN LO | 01-N-57 | | |
| THE VOTING IS OVER (TRANSITION OF PRESIDENTIAL ADMINISTRATION). | AFTER | LONG NE | 06-M-62 | | |
| CHARACTERISTICS AND LEGISLATIVE OVERSIGHT OF ADMINISTRATION. | COMMITTEE | RIDLEY JF | 10-F-64 | | |
| POLICY MAKING BY GOVERNMENT AGENCIES (PUBLIC ADMINISTRATION, BUREAUCRACY). | | ROYER WM | 04-A-60 | | |
| AS IN THE RHODESIAS. | ADOPTION OF POLITICAL STYLES BY AFRICAN POLITICI | SCARRITT J | 10-F-66 | | |
| CONGRESSIONAL RESISTANCE TO REFORM-- | ADOPTS A CODE FOR INVESTIGATING COMMITTEES. | HEUREL EJ | 01-N-57 | | |
| 1958 GENEVA CONFERENCE OF EXPERTS (TECHNICAL ADVISORS). | SCIENTISTS AS NEGOTIATORS-- | STICKLA D | 08-N-64 | | |
| TIVE RELATIONS. | AEC AND CONGRESS-- A STUDY IN EXECUTIVE - LEGISL | NIERBURG HL | 06-M-62 | | |
| IZATION AND CHILDREN'S ATTITUDES). | AFFAIR IN PERSPECTIVE (GERMANY). | GIMMEL J | 09-A-65 | | |
| T PERCEPTION OF THE DEVELOPMENT PROCESS IN THE | AFFILIATION AND IMAGERY ACROSS TWO RELATED GENER | DOUGLE RW | 06-A-62 | | |
| POLITICAL | THE AGE FACTOR IN THE 1958 CONGRESSIONAL ELECTIONS. | CARLISLE D | 08-N-64 | | |
| THE | AGENCIES (PUBLIC ADMINISTRATION, BUREAUCRACY). | WALKER DH | 04-F-60 | | |
| POLICY MAKING BY GOVERNMENT | AGENCIES AND AN APPROPRIATIONS SUBCOMMITTEE-- | HOYER WM | 04-A-60 | | |
| COMPARATIVE STUDY OF BUDGET STRATEGIES. | AGREEMENT BETWEEN JUSTICES OR LEGISLATORS. | SHARKANS I | 09-A-65 | | |
| INDICES OF PAIRWISE | THE COMMODITY CREDIT | MACRAE D | 10-F-66 | | |
| RPORTATION AND THE 1948 PRESIDENTIAL ELECTION (AGRICULTURE). | | WILLIAMS O | 01-A-57 | | |
| ICE SUPPORT POLICY AND THE MIDWEST FARM VOTE (AGRICULTURE). | | GILPATRI T | 03-N-59 | | |
| THE ROLE OF THE CONGRESSIONAL SUBCOMMITTEE (AGRICULTURE APPROPRIATIONS COMMITTEE). | | JONES CO | 06-N-62 | | |
| FOREIGN | AID AND THE NATIONAL INTEREST. | PACKENHA R | 10-M-66 | | |
| CY COMMITTEES AND INCREMENTALISM-- | INTERA | RIPLEY RB | 08-M-64 | | |
| THE CASE OF | ALEXANDER HAMILTON AND THE AMERICAN TRADITION (T | LIVINGST J | 01-N-57 | | |
| THEORY). | ALTERNATION REFERENDA. | JOYNER C | 02-M-58 | | |
| AUSTRALIAN POLITICS AND CONSTITUTION | AMBIGUITY(UNITED STATES CIVIL SERVICE, GOVERNME | NELSON DM | 02-F-58 | | |
| DER THE MATCH ACT AND THE PROBLEM OF STATUTORY | AMENMENT, PRESIDENTIAL ELECTION). | GOLDMAN RM | 02-F-58 | | |
| L FOR ELECTORAL COLLEGE REFORM(CONSTITUTIONAL | AMERICA. | POLIT ANDERSON C | 05-N-61 | | |
| CS AND DEVELOPMENT POLICY IN CENTRAL (LATIN) | TAXES ON THE #EXTERNAL# SEC | KLING M | 03-M-59 | | |
| OR-- AN INDEX OF POLITICAL BEHAVIOR IN LATIN - | AMERICAN) POLITICAL BEHAVIOR. | MILLER WE | 02-A-58 | | |
| THE SOCIO-ECONOMIC ANALYSIS OF (| AMERICAN CITIES. | DYE TR | 08-N-64 | | |
| ION-- CONDITIONS ASSOCIATED WITH ANNEXATION IN | AMERICAN LAW). | SCHUBERT G | 02-F-58 | | |
| E PUBLIC INTEREST IN JUDICIAL DECISION-MAKING(| AMERICAN POLITICS-- A CLARIFIED VIEW. | SCHLESIN J | 01-M-57 | | |
| LAWYERS AND | AMERICAN PRESIDENCY. | SIGEL RS | 10-F-66 | | |
| IMAGE OF THE | AMERICAN STATE LEGISLATORS. | EULAU M | 03-M-59 | | |
| THE POLITICAL SOCIALIZATION OF | AMERICAN TRADITION (THEORY). | LIVINGST J | 01-N-57 | | |
| ALEXANDER HAMILTON AND THE | AMERICAN UNIONISM (INTEREST GROUPS). | SEIOLER M | 05-A-61 | | |
| THE SOCIALIST PARTY AND | AMPHIBIAL STATE (THEORY). | SMITH JM | 01-M-57 | | |
| FREEDOM AND AUTHORITY IN THE | ANALYZING LEGISLATIVE BEHAVIOR. | RELKNAP GM | 02-N-58 | | |
| A METHOD FOR | ANCIENT INDIAN POLITICAL THOUGHT(BUDDHISM). | ALBINSKI H | 02-F-58 | | |
| THE PLACE OF THE EMPEROR ASOKA IN | ANNEXATION IN AMERICAN CITIES. | DYE TR | 08-N-64 | | |
| TICAL INTEGRATION-- | CONDITIONS ASSOCIATED WITH | | | | |

Figure 11. KWIC computer output: sample keyword listing from the cumulative index to the Midwest Journal of Political Science.

FORTRAN AND ALGOL BEAR TO THE PERFORMANCE OF ARITHMETIC, THAT IS, A USER ORIENTED SYMBOL M 409
 FORTRAN AND ALGOL USE THE FORMULA TO EXPRESS IN A SINGLE LINGUISTIC STRUCTURE AN ARBITRARY 411
 FORTRAN ASSIGNMENTS ARE A SUBSET. WHEN THIS ASSIGNMENT-STATEMENT ABILITY IS COUPLED WITH 411
 FORTRAN BY PRECOMPILERS. "A. WERTZ H J PROGRAM-WRITING PROGRAMS, EXTENDING 650
 FORTRAN CAN MAKE EFFICIENT USE OF SUCH PRECOMPILERS TO GENERATE SPECIAL-PURPOSE PROGRAMS C 625
 FORTRAN COMPUTER PROGRAMS FOR INFORMATION RETRIEVAL (VINSOHALER), AND ANOTHER ANNOUNCEMEN 625
 FORTRAN II FOR THE IBM 1620. TWO MODEL VARIATIONS, INPUT AND OUTPUT EXAMPLES, AND SOME PO 564
 FORTRAN IV CONTAINS SUCH EXTENDED FEATURES AS MIXED MODE EXPRESSIONS, GENERALIZED SUBSCRIP 787
 FORTRAN IV HAS BEEN IMPLEMENTED UNDER THE CONVERSATIONAL COMPILER SYSTEM (CCS), A GROUP OF 668
 FORTRAN IV IN A FORM COMPATIBLE WITH THE CDC 3400 SYSTEM CURRENTLY INSTALLED AT NORTHWESTE 394
 FORTRAN IV PROGRAM FOR THE CDC 6400 COMPUTER WHICH COMPUTES TWO-AND THREE-STAGE LEAST SQUA 556
 FORTRAN IV PROGRAMMING AND COMPUTING. " GOLDEN JAMES T 019
 FORTRAN IV REFERENCE MANUAL. " COM-SHARE, INC 666
 FORTRAN LANGUAGE. BOTH EDGE-DETECTION AND TOPOLOGICAL PREPROCESSING ARE COUPLED WITH CLA 769
 FORTRAN PROGRAM FOR THE CDC 3400. "S 567
 FORTRAN PROGRAMMING. " HARVILL JOHN B BASIC 021
 FORTRAN PROGRAMMING. " RULE W P INTRODUCTION TO 047
 FORTRAN PROGRAMS DEPENDING ON ONLY A LIMITED SET OF #PRIMITIVE# MACHINE LANGUAGE SUBROUT 630
 FORTRAN SYNTAX ANALYSIS. " 769

Figure 12. KWIC output using NOWRAP feature.

The decision as to what is or is not to be considered as a keyword is made in three ways. The computer can be instructed either to refer to a list of keywords prepared in advance by the researcher (KEY), or to a list that is not to be considered as keywords (stopword list), or can be instructed to index on every term in the file (NULL).

In the first case, the computer looks at every word contained in the level of information it is to search, compares it with its stored list of keywords, and selects matches for indexing. The process operates in a comparable way when a list of "stopwords" is used: the computer includes the word in the index only when it does not appear in the list. Sample stopwords are "a", "an", "of", "the", etc. but this list can be quite distinctive when applied to specific files. For example, a file of computer-related literature would have "computer" as a stopword since it is assumed that all entries would fall into this classification.

A KWIC index is also known as a "permuted" index, for an entry will appear as many times as the number of keywords it contains. The sample output in Figure 10 from the Midwest Journal file contained almost 200 articles, each of which was indexed an average of five times in a listing almost 1000 entries long.

When the KWIC (or AUTHOR) forms of output are selected, the only indication of the source of the entry is the reference code. This may, or may not, depending on the users' needs, require the generation of a separate "bibliographic" form of output in which the complete entry is listed in order of reference code. This is one form of KWOC output, with the reference code generated by the EDIT program used as the level of information (0) to be searched. This points up the fact that KWIC indexes are generally "double-entry" indexes; a KWOC index, on the other hand, is a "single-entry" index,

for it prints out (upon user option) as much of the complete entry as is desired for each appearance of a keyword.

Figure 13 reproduces a portion of output from a title index to the Vogelback Computing Center library. The index term is at the left of the page and each item indexed under that term is printed below. In this case, only author, title, and source information is printed. In other indexes to the same file, abstracts, coding categories, and shelf locations were printed.

Figure 14 shows an index to the same file where authors (level one information) were the designated search level. Another form of author index is a "short-form" index produced by the use of the *AUTHOR card. In this index, only the author's name and the reference code associated with that entry is listed (Figure 15).

The INDEX routines provide the user with the capability to choose which levels of information are to be searched and which levels are to be carried along as input to the PRINT routines (see p. 43). In addition, a STATISTICS option is available for counting and printing the total number of occurrences in a file of a key or stopword; the FREQUENCY option will print a listing of all words and their frequency within a given entry which will be used as indexing terms in KWOC output. This option is very useful when working with a new or large file; the frequency list can be printed and checked for desirability of indexing terms without printing the index itself. In this way, additional stopwords may be added to produce a final output pertinent to the user's needs.

SIMULATION

588

GAFARIAN A V ET AL

STATISTICAL APPROACH FOR VALIDATING SIMULATION MODELS BY
COMPARISON WITH OPERATIONAL SYSTEM - ILLUSTRATED FOR
TRAFFIC FLOW.
SYSTEM DEVELOPMENT CORPORATION, SANTA MONICA, CALIFORNIA.
FEBRUARY 1966. AD 632 478

654

GREENBERGER M

A NEW METHODOLOGY FOR COMPUTERS SIMULATION.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE,
MASSACHUSETTS. 1964. AD 609 288

672

SMITH D L

MODELS AND DATA STRUCTURES FOR DIGITAL LOGIC SIMULATION.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE
MASSACHUSETTS. AUGUST 1966. AD 637 192

753

ROME B K

ROME S C

HAYTHORN W W

INFORMATION SYSTEMS SIMULATION AND MODELLING TECHNIQUES,
SESSION 7, FIRST CONGRESS ON THE INFORMATION SYSTEM
SCIENCES

THE MITRE CORPORATION, BEDFORD, MASSACHUSETTS, DECEMBER 1963
AD 426 985

787

CONGER C R

THE SIMULATION AND EVALUATION OF INFORMATION RETRIEVAL
SYSTEMS

HRB - SINGER, INC., SCIENCE PARK, STATE COLLEGE,
PENNSYLVANIA, APRIL 1965, AD 464 619

SIR

652

RAPHAEL R

SIR, A COMPUTER PROGRAM FOR SEMANTIC INFORMATION RETRIEVAL.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE,
MASSACHUSETTS. JUNE 1964 AD 608 499

SNOBOL3

537

FORTE A

SNOBOL3 PRIMER.

THE MIT PRESS 1967, CAMBRIDGE, MASS.

SNOBOL4

631

GRISWOLD R E

POAGE J F

POLONSKY I P

PRELIMINARY REPORT ON THE SNOBOL4 PROGRAMMING LANGUAGE.
INSTITUTE FOR DEFENSE ANALYSES, PRINCETON, NEW JERSEY.
MARCH 1 1968.

IVIE E L
 IVIE E L
 SEARCH PROCEDURES BASED ON MEASURES OF RELATEDNESS BETWEEN
 DOCUMENTS.
 MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE,
 MASSACHUSETTS. JUNE 1966. AD 636 275

JAHODA G
 JAHODA G
 ANALYSIS OF CASE HISTORIES OF PERSONAL-INDEX USE
 FLORIDA STATE UNIVERSITY, TALLAHASSEE, FLORIDA, JULY 1966.
 AD 643 869

JANDA K
 JANDA K
 CUMULATIVE INDEX TO THE AMERICAN POLITICAL SCIENCE REVIEW. =
 NORTHWESTERN UNIVERSITY PRESS, 1964.
 COMPUTING CENTER LIBRARY.

JANDA K
 DATA PROCESSING - APPLICATIONS TO POLITICAL RESEARCH. =
 NORTHWESTERN UNIVERSITY PRESS, 1965.
 COMPUTING CENTER LIBRARY.

JANDA K
 A MICROFILM AND COMPUTER SYSTEM FOR ANALYZING COMPARATIVE
 POLITICS LITERATURE.
 ICPP PROJECT, NORTHWESTERN UNIVERSITY, 1967.
 COMPUTING CENTER LIBRARY.

JANDA K (ED)
 JANDA K (ED)
 ADVANCES IN INFORMATION RETRIEVAL IN THE SOCIAL SCIENCES,
 PART 1.
 AMERICAN BEHAVIORAL SCIENTIST, JANUARY, 1967.

JANDA K (ED)
 ADVANCES IN INFORMATION RETRIEVAL IN THE SOCIAL SCIENCES,
 PART 2.
 AMERICAN BEHAVIORAL SCIENTIST, FEBRUARY, 1967.

JOHNSON C H
 JOHNSON C H
 DATA PROCESSING. =
 NATIONAL MACHINE ACCOUNTANTS ASSOCIATION, 1959.
 COMPUTING CENTER LIBRARY.

RUBINOFF M
RUBINOFF M ET AL
RUBINOFF M ETAL
RUDIN B D
RULE W P
RUSSO F J
SAILER A J
SAKODA J M
SALTZER J M

SALZMAN R M ET AL
SAMUEL A L
SANCHEZ A
SATTERLY G T JR
SAVITT D
SAYER J S
SCHEFFLER F L
SCHERR A L
SCHMIDT R N
SCHULTZ L
SCHUMAN, S A
SCHWEPPE E J
SELBY S M
SHENTON L R

SHER I H
SHINDLE W E

SIMMONS R
SINGER A J
SINNETT J D
SKOGAN W
SMALLWOOD R O ETAL
SMITH A A
SMITH C
SMITH D L
SNELL J L
SNOODEY S R

730
595
739
741
047
649
761
048
012
659
663
604
664
624
049
722
764
704
648
050
762
414
033
688
523
524
699
586
587
776
680
543
475
686
657
774
672
029
755

SWEEN J
TAFT T D
TAYLOR R S ETAL
TEITELMAN W

THOMAS G B JR
THOMPSON C W N

THOMPSON F B
THOMPSON G L
THORPIE R J
TIERNEY J M
TOMLINSON H
TONGE F M
TORTOLINI A T
TRIST W E
TROOP R
TRUESWELL R W
TWITE M J
UNIV OF GEORGIA
UTTAL W
VAN HORN E C JR
VAREHA A L, JR
VINSONHALER J F
VON FOERSTER H
VORHAUS A H
WALLACE E M

WANTMAN M E
WARD J E
WARSHALL S
WAY W
WEAST R C
WEIZENBAUM J
WERNER D J
WERNER O
WERTZ H J

557
567
745
689
602
673
056
608
619
681
029
699
775
718
770
529
607
722
581
754
058
759
651
628
551
725
574
549
617
712
650
580
40
714
688
630
582
059
625

Figure 15. *AUTHOR (short-form) Index

INDEX Control Cards

All INDEX control cards are punched free-field, i.e., no column requirements. All instructions contained within brackets are optional. A period must be the last character on the *WORDS, *KWIC, and *KWOC cards. Each desired KWIC or KWOC index requires a set of INDEX and PRINT control cards (see sample job deck structure, p.46). Multiple indexes may be produced in one computer run by supplying multiple sets of INDEX and PRINT instruction cards, one set per printed output.

*MASTER = n

This instruction is used only when an existing (previously created and saved) file of information is to be used as input to the indexing routines, i.e., not as part of an EDIT and INDEX operation. n may be any unit from 1 to 5 and must be referenced on a system REQUEST card.

*INDEX (required)

To call in indexing routines

*WORDS, name of index, CARDS (n_1, n_2, \dots, n_m), [NULL], [STATISTICS] .

where name of index is user defined;
can be any simple word or acronym meaningful to the user. This name must be repeated on the *KWIC or *KWOC card associated with this index. The WORDS card defines the levels of information to be searched ($CARDS(n_1, \dots, n_m)$). The wordlist, if supplied, will be matched against terms contained only within these information levels.

The NULL option indicates that a word list is not supplied; therefore, all words will be used as indexing terms. STATISTICS will produce a frequency count of the number of times each word in the user-supplied word lists was encountered in the file.

Word list, either "key" or "stop", follows the *WORDS card. Words are punched free-field, separated by commas, with the last word ending in a period.

*KWIC, name, [KEY], CARDS (n₁, n₂...n_m), [NOWRAP] .

The *KWIC card indicates that output is to be in KWIC format. CARDS (n₁...n_m) specify which levels of information are to be printed. "name" must be the same as specified on the *WORDS card.

KEY indicates that a keyword list of indexing terms has been supplied and is to be used.

NOWRAP indicates output is nowrap format, i.e., keyword on left of page followed by context. If not specified, output will be "standard" KWIC, i.e., keyword running down vertical column of page, surrounded on both sides by context. (See Figure 10, p. 31)

*KWOC, name, [KEY], CARDS (n₁, n₂,...n_m), [FREQUENCY] .

Indicates output is to be in KWOC format.

"name" must be the same as that specified

on the WORDS card.

CARDS ($n_1, n_2 \dots n_m$) specifies which levels of information are to be carried into the PRINT program.

Note that if printing of a reference code is desired, the "0" level must be indicated.

KEY indicates that a user-supplied key-word list is to be used as index terms. If not specified, a stopword or null list is assumed.

The FREQUENCY option will produce a list of the words that will be used for indexing and their frequency within each entry in the file. The frequency option does not accumulate a count over the entire file.

*AUTHOR

To produce a short-form alphabetized listing of authors (level one information) and the reference code associated with each entry. Output for single authors is triple column; corporate authors appear on a separate list in single column format. When using the *AUTHOR option, a *PRINT card must be included in the set even though print options are not available.

*END

To return control to the executive overlay.

PRINT Routines

Various output options are available to the user in conjunction with the indexing routines. A call to the PRINT overlay is required for each printed index.

PRINT Control Cards

*PRINT

To call in the print routines.

Print Option Card KWOC Output: (All words enclosed within brackets are optional) A period must follow the last specified option.

OK.

Indicates default print conditions.

Output will be:

double column

no indentation

all levels of information will be
printed

reference code is not printed

line count per page = 60

spacing at top of page = 10

spacing on first page = 20

DOUBLE [COLUMN],

SINGLE [COLUMN],

CARDS (n_1 , n_2 , n_m),

To specify which levels of information
are to be printed

REFERENCE [CODE] ,

To print reference code

[NUMBER OF] LINES = n ,

To specify number of lines per page

[SPACING AT] TOP = n ,

To specify number of lines to be dropped
at top of page before printing

[ADDITIONAL AT] BEGINNING = n,

To specify number of additional lines to be dropped on first page of printed index. This space may be used for title insertion.

OUTPUT = u,

where u must be either 2, 4 or 5;
u is a physical tape unit that output will be written on. This option is normally used when special printing is desired. The output tape can be resubmitted at a later time for printing using a COPY routine. A REQUEST card must be supplied.

COPY = n,

n is the number of times the output is to be written on the user supplied output tape. This is useful when multiple copies of an index are desired.

INDENTATION

Output entries are indented (within entry).

*END

To return control to main overlay.

Print Control Card for KWIC Output: (All words enclosed within brackets are optional) A period must follow the last specified option.

OK.

Indicates default print options desired.

Output will be:

lines per page = 60

spacing at top of page = 10

additional spacing on first page = 20

Options:

[NUMBER OF] LINES = n,

[SPACING AT] TOP = n,

[ADDITIONAL AT] BEGINNING = n,

OUTPUT = u,

COPIES = n,

*END

To return control to executive overlay.

When using the *AUTHOR option, the only PRINT cards needed are:

*PRINT

*END

TO PRODUCE AUTHOR AND TITLE INDEXES IN KWOC FORMAT FROM CARD INPUT:

(Note that although an EDIT of the input data must be performed before calling the INDEX routines, this file will not be saved for any subsequent processing since a disk scratch file is automatically assigned to the "master" file if a tape assignment using the REQUEST card is not supplied.)

VCC System
Cards

```
VLORR,CHCM1235-9999,CM71200,T1000.  
LIBRARY(TRIAL3C)  
LGO.  
7-8-9 End of Record
```

Create a file:

```
*EDIT  
NEW MASTER = 1.  
*DATA  
*INSERT,  
data cards  
*END
```

Generate an
unformatted
author index

```
*INDEX  
*WORDS,AUTHOR,CARDS(1).  
user-supplied stopword list  
*KWOC,AUTHOR,CARDS(1,2,3).  
*END
```

Sort and print
formatted author
index

```
*PRINT  
OK.  
*END
```

Generate unformatted
title index; print
list of words that
each entry will be
indexed under, together
with frequency of occurrence.

```
*INDEX  
*WORDS,TITLE,CARDS(2).  
stopword list  
*KWOC,TITLE,CARDS(1,2,3),FREQ.  
*END
```

Print in single
column format title
index, listing author
and title only.

```
*PRINT  
SINGLE COLUMN, CARDS(1,2).  
*END
```

6-7-8-9 End of Information

Figure 16.

References

1. Aagaard, James S. and Roger J. Schaid, "Research on Advanced Telemetry Techniques: Bibliographic Data Processor", Aerial Measurements Laboratory, Northwestern University, Contract AF33(615)-1945, December 28, 1964.
2. Aagaard, James S., "BIDAP", American Behavioral Scientist, 10, February, 1967.
3. Borman, L., and Dillaman, D., "TRIAL for the CDC 3400", Vogelback Computing Center, Northwestern University, July 25, 1967.
4. Janda, Kenneth, "Keyword Indexes for the Behavioral Sciences", American Behavioral Scientist, 7, June, 1964.
5. Janda, K., and Milbrath, Lester W., "Computer Applications to Abstraction, Storage, and Recovery of Propositions from Political Science Literature", Paper presented at the 1964 Annual Meeting of the American Political Science Association.
6. Janda, K., (ed), Cumulative Index to the American Political Science Review, Volumes 1-57, 1906-1963, Evanston: Northwestern University Press, 1964.
7. Janda, K., Data Processing: Applications to Political Research, Evanston: Northwestern University Press, 1965.
8. Janda, K., "Information Retrieval: Applications to Bibliographies on International and Comparative Politics", Paper presented at the Computers and the Policy-Making Community Institute, Lawrence Radiation Laboratory, Livermore, California, April 4-15, 1966.
9. Janda, K., and Tetzlaff, William H., "TRIAL: A Computer Technique for Retrieving Information from Abstracts of Literature", Behavioral Science, 11, November, 1966.
10. Janda, K., and Rader, Gary, "Selective Dissemination of Information: A Progress Report from Northwestern University", American Behavioral Scientist, 10, January, 1967.
11. Janda, K., Information Retrieval: Applications to Political Science, Indianapolis and New York: The Bobbs-Merrill Company, Inc., 1968.
12. Tetzlaff, William H., "TRIAL: Technique for Retrieving Information from Abstracts of Literature - a Program for the IBM 709/90/94", Paper presented at the SHARE XXVI meeting, Applied Management Sciences Project Committee, San Diego, California, February 28-March 4, 1966.

TRIAL Internal Record Format (to be used when altering lines of an existing entry (ALTER))

| | |
|------------|---|
| Col. 1-60 | New line of input data |
| Col. 61 | Level number |
| Col. 62-66 | Line number to be altered. EDIT LIST must be referenced. |
| Col. 67-68 | Repeating group identifier, if any. EDIT LIST must be referenced. |
| Col. 69-80 | Reference information as punched in col. 61-72 of original data. |

| KEY WORDS | LINK A | | LINK B | | LINK C | |
|---|--------|----|--------|----|--------|----|
| | ROLE | WT | ROLE | WT | ROLE | WT |
| <p>Information storage and retrieval</p> <p>Computer programs</p> <p>Programming Language</p> <p>Large files</p> <p>FORTRAN</p> <p>Natural language data</p> <p>Text manipulation</p> <p>Social sciences data</p> <p>Indexing</p> | | | | | | |